Cellular and Molecular Biology™ **52**, N°6,47-52
DOI 10.1170/T735

ISSN 1165-158X
2006 Cell. Mol. Biol.™

AN AUTOMATIC SYSTEM FOR THE ANALYSIS OF INTERCELLULAR COMMUNICATION AND EARLY CARCINOGENESIS

N. LOMENIE[2✉], F. RICHARD[1], E. DUSCH[1,2], D. SEGRETAIN[3]
[1✉]MAP5 lab (UMR CNRS 8145), Dpt .of Mathematics and
Computer Science, University René Descartes,
45 rue des Saints-Pères, 75006 Paris, France
Fax: *+33 1 44 55 35 35*; *E-mail*: *Nicolas.Lomenie@math-info.univ-paris5.fr*
[2]SIP-CRIP5 lab, Dpt. of Mathematics and Computer Science, University René Descartes
[3,4]INSERM EMI 00-09, Dpt. of Biology, University René Descartes

Received March 11th, 2005; *Accepted April 20th, 2005*; *Published May 30th, 2006*

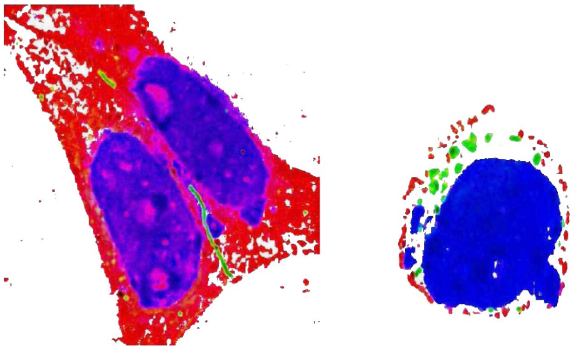
Abstract – Some recent works on intercellular communication pointed out an impaired trafficking of Cx43 proteins in early carcinogenesis. In collaboration with biologists, we propose an automatic system for the analysis of spatial protein configurations within cells at early tumor stages. This system is an essential step towards the future development of a computer-aided diagnosis tool and the statistical validation of biological hypotheses about Cx43 expressions and configurations during tumorogenesis. The proposed system contains two dependent part: a segmentation part in which the cell structures of interest are automatically located on images and a characterization part in which some spatial features are computed for the classification of cells. Using immunofluorescent images of cells, the nucleus, cytoplasm and proteins structures within the cell are extracted. Then, some spatial features are computed to characterize spatial configurations of the proteins with regard to the nucleus and cytoplasm areas in the image. Last, the 3D cell images are classified into pathogenic or viable classes. The system has been quantitatively evaluated over 60 cell images acquired by a deconvolution high-resolution microscope and whose ground truth has been manually given by a biologist expert. As a perspective, a 3D spatial reasoning and visualization module is currently under development.

Key words: gap channel junctions, connexin, intercellular communication, carcinogenesis, immunofluorescent microscopy, image analysis, image processing, image segmentation, image classification, 3D cell visualization, spatial analysis.

INTRODUCTION

Since its emergence from the pioneer work of Lowenstein (1), the biological hypothesis that the cell-to-cell communication could be involved in the control of the cell growth has been confirmed by various experiments (see (2) for a review). Cells communicate to their neighboring cells thanks to gap junction channels (GJC) which allow the direct transfer of signal molecules across cell membranes. These channels are formed of proteins called connexins (Cx), among which Cx43 is the most representative in mammalians. Recently, impaired GJC and Cx expressions have been demonstrated *in vivo* in some human tumors (see (3) for a review). Some experiments also suggested that genes coding for Cx could be tumor-suppressor (4) and that Cx could be used in pharmacological treatments of cancers(5). However, designing such a clinical application still requires a better understanding of Cx expression during carcinogenesis.

In the last few years, Segretain and colleagues focused their researches on early stages of the testicular carcinogenesis. Using transgenic mouse model, they obtained initial evidence that Cx43 in and scatter within the cells (fig. 1.b). Some other Leydig cells were aberrantly localized in cells at early tumor stages (6). These results were obtained thanks to a deconvolution high-resolution microscope (DVHR), which generates sharp cell images at a sub-cellular resolution (67nm) and allows to observe precisely various cell structures of interest. For the acquisition of the DVHR images, cell structures were enhanced with immunofluorescent markers: the blue marker diamidino-phenylindole (DAPI) for nuclei, the green marker fluorescein isothiocynanate (FITC) for Cx43 and the red marker rhodamine (RHOD) for cell cytoplasms. The enhanced structures were made visible separately in the RGB color channels of images. Thanks to a piezoelectric translator, the microscope also collected 2D color slices of cells at 0.2µm Z-intervals. In fig. 1, some typical examples of image slices generated by the DVHR microscope illustrate cases when the Cx expression in the cell is normal and the trafficking of Cx is impaired due to the cancer. As it can be seen, Cx are normally aligned all along cell membranes (fig. 1.a), but, at early carcinogenesis, Cx migrate towards the nuclei illustrations may be found in (6).



(a) (b)

Fig; **1** Some examples of 2D cell slices obtained by Segretain et al. with the DVHR microscope (1). The nuclei, the cytoplasms and the Cx43 within the cells can be observed in the color channels of images: blue, red and green, respectively. Images show cells of a control case (a) and of a pathological case (b) (early carcinogenesis).

Observations of Segretain et al. (6) suggested that the spatial arrangements of Cx43 relative to cell membrane and nucleus positions could characterized the pathological cell states of carcinogenesis. The long-term goals of our work are (a) to validate statistically biologist assumptions and (b) to build a tumor detection system which could be used as a tool for computer-aided diagnosis. This paper presents the preliminary step towards the realization of these goals. It consists in designing and evaluating an automatic system which analyzes spatial arrangements of Cx43 within the cells. This system is mainly composed of two dependent parts, a segmentation part in which the cell structures of interest are automatically localized and a characterization part in which some spatial features are computed with the aim to classify cell states (e.g. control/pathological states).

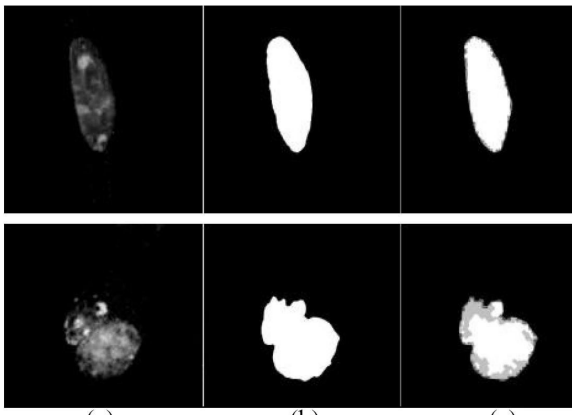
Section 2 describes the various parts of the analysis systems. In Section 3, based on a test set of images annotated by an expert, the results of an evaluation of the system are given and some improvements of the system are also presented.

MATERIALS AND METHODS

The three channels of images generated by the DVHR microscope contain complementary information about cellular structures and none of them are sufficient to obtain an exhaustive cell description. It is thus necessary to process all of the channels and to combine results. Besides, due to the nature of the acquisition, the image appearance is quite different from a channel to another. Hence segmentation techniques have to be specific to each channel. In what follows, we focus successively on the segmentation of cell structures in each image channel directly converted into grey- level images.

DAPI image segmentation

By visual inspection of DAPI images, the nuclei can be easily located. Since these images are well-contrasted, it is also possible to have a good initial estimate of nucleus positions using a thresholding operation (see fig. 2.a).



(a) (b) (c)

Fig. **2** **(a)** DAPI image of a nucleus (b) ground truth or gold standard obtained by a biologist expert (c) algorithmic segmentation of the nucleus region with the gray regions corresponding to the difference with the ground truth in b).

However, the choice of the threshold value for nucleus segmentation is a critical implementation issue. Due to DAPI image variations, there is no threshold which is suitable for all images. Instead, a strategy must be defined for computing automatically a suitable threshold for each image. In DAPI images, gray-level values in nucleus regions are high and variable, as opposed to those in the background which are quite low and uniform. As a consequence, gray-level histograms of the DAPI images always show bimodal distributions (see fig. 3).

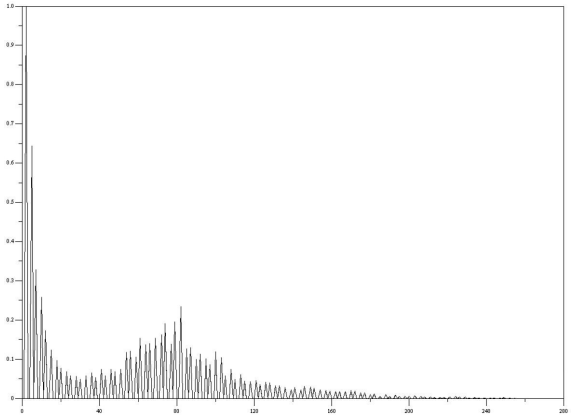
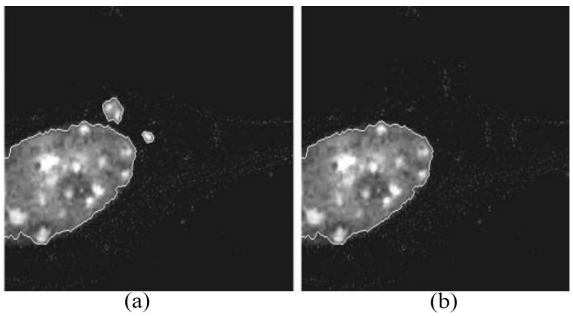


Fig. **3** Bimodal histogram of the DAPI nucleus image of fig. 2.

Hence, gray-level values which separate the two modes are relevant candidates for thresholding images. Such values can be optimally computed using a *k-means* clustering method (7).

The segmentation obtained using thresholds cannot be perfect. Indeed, the automatic thresholding operator does not provide a filtered, smooth shape of the nucleus area. The marker used for the nucleus also marks some small regions in the cell outside the nucleus (see fig. 4.a).



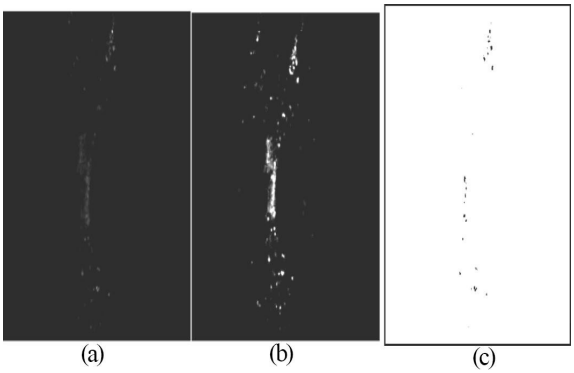
(a) (b)

Fig. 4 Segmentation of the DAPI nucleus image by automatic thresholding (a) processed by morphological operators (b).

To fix the problem, morphological opening and closing operators are respectively used to delete small particles and to improve the under-segmentation of nucleus (8). The structuring element is a circle with a radius adapted to the resolution.

FITC image segmentation

As protein images are well-contrasted, a simple contrast enhancement filter followed by an automatic Otsu-type thresholding technique (9) is sufficient to extract protein structures (see fig. 5). This basic operation can be improved by using a wavelet-based multi-resolution analysis (10): as aAn automatic system for the analysis of intercellular communication and early carcinogenesis matter of fact, the multiscale correlation of the filtered wavelet coefficients allows to enhance multiscale peaks due to spots while reducing noise by combining information coming from different levels of resolution. This method is more robust to the variability of biological images and to the high level of noise present in typical immunomicroscopic images. To filter the shape of the obtained cytoplasm segmentation, the same opening and closing morphological operators as in the case of the nucleus segmentation are applied. Besides, the complementary segmentations of RHOD, DAPI and FITC images permit to infer a segmentation of the cell region: the cell mask is obtained as the union of the cytoplasm region (RHOD segmentation), the nucleus region (DAPI segmentation) and the protein regions (FITC segmentation). of of the preliminary cell segmentation mentioned in the previous parts (proteins and nucleus) to obtain a complete. An example of cell segmentation is shown in fig. 8. Some examples of complete segmentations of cell structures are also shown on fig. 9.

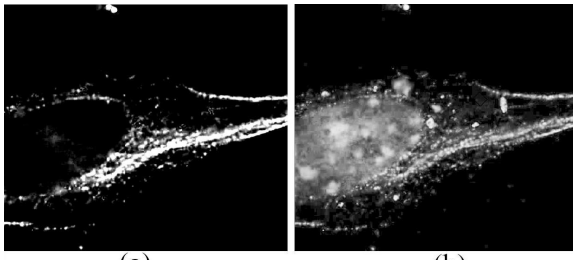


(a) (b) (c)

Fig. 5 (a) FITC proteins image (b) Image processed by a contrast enhancement operator (c) Extracted proteins in the image (a).

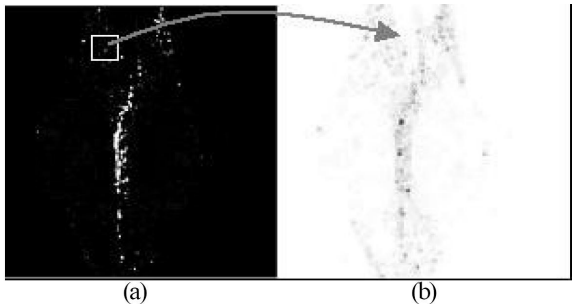
RHOD image segmentation

Because of the granular aspect of RHOD images (see fig. 6), thresholding directly these images is not a relevant segmentation technique. Indeed, such a technique would result in an incomplete segmentation of cytoplasm regions strewn with holes corresponding to dark image areas. Hence, instead of the raw RHOD image, a feature image is derived from the raw image by computing the local variance of the gray levels within a window of size *n* centered around each pixel (see fig. 7). The feature image is then segmented using the same thresholding technique as in the case of DAPI images.



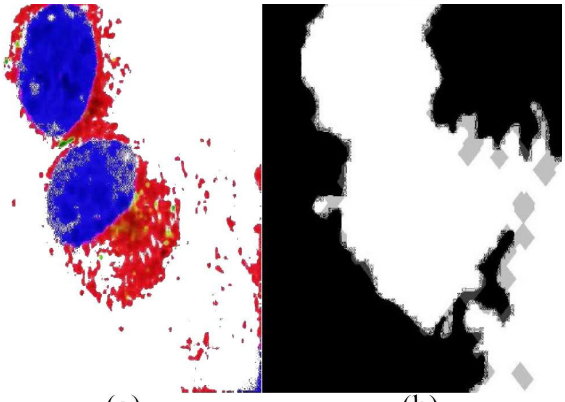
(a) (b)

Fig. 6 An example of (a) a RH0D image and of (b) its associated cytoplasm area (union of the segmented nucleus area with the RHOD marked area).



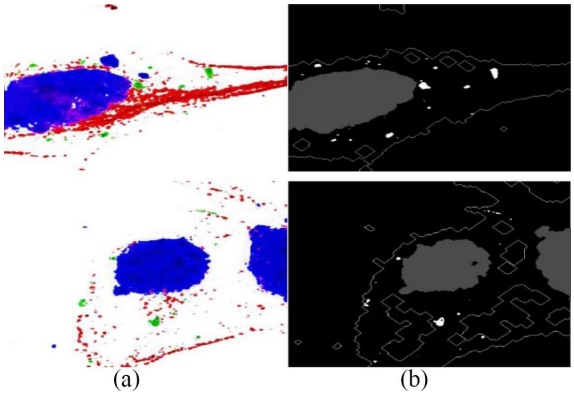
(a) (b)

Fig. 7 An example of a RH0D image (a) and of its associated rugosity measure image (b) computed with a mobile window of fixed size.



(a) (b)

Fig. 8 (a) DVHR images and (b) the overlay of the computed cell segmentation and the ground truth (gray regions correspond to differences between the computed segmentation and the ground truth).



(a) (b)

Fig. 9 (a) Immunofluorescent images (b) with their corresponding cellular mask composed of the union of the three specific segmentation results : white contour for the cell membrane, gray region for the nucleus and white spots for the proteins.

RESULTS

Image Segmentation

A database of images has been built up with 60 different slices taken from a test set composed of 10 control samples and 10 pathological samples. Each sample contains one or two cells. Three successive slices were taken from each sample of the test set. Besides, the segmentation results obtained on control and pathological cells were equivalent. Each image slice has been manually segmented by a biologist expert. Due to a globally low signal-to-noise ratio and to some labelling problems for a few images, the segmentation results are actually evaluated on a set of 33 image slices taken from 11 various cells (see fig. 10).

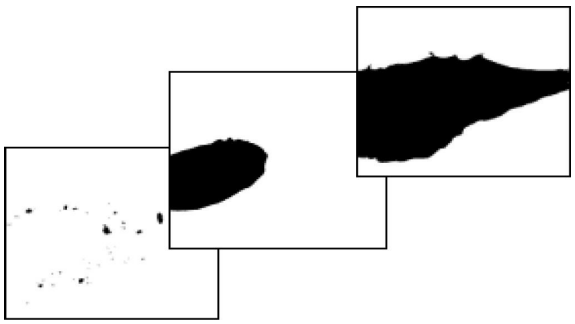


Fig. 10 Masks for the ground truth. These masks were manually segmented by a biologist.

Although these results are quantitatively good, it is worth giving some comments about achieved segmentations and some difficulties which arise. In images where several cells are present, membranes of the various cells are often in contact. Consequently, the segmentation technique may not distinguish between neighboring cells and may merge distinct cells into a single structure (see an illustration in fig. 11.a and 11.b).

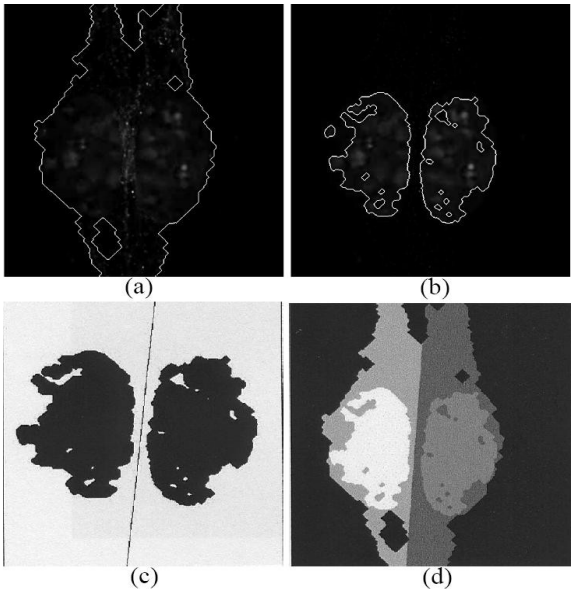


Fig. 11 (a) The segmentation processing merges neighboring cells (b) The segmentation of the corresponding nuclei (c) Line separating the nuclei regions (d) The correction of the segmentation in (a).

The segmentation evaluation is performed by the means of the Dice Similarity Coefficient (DSC) (11). This criterion is a measure of the spatial overlap between a region A (ground truth given by the expert) and a region B (computed automatically):

$$DSC(A,B) = 2 * (A \cap B)/(|A|+|B|)$$

The value of a DSC ranges from 0 (no spatial overlap) to 1 (perfect overlap). As recommended by Zijdenbos in (12), an overlap is good whenever DSC is above 0.7. Results of the evaluation on the test image set are given for each structure on table 1. For nuclei and cells, averages of DSC are above 0.85 and standard deviations are very low, showing that the techniques proposed for the segmentation of these structures are reliable. The segmentation results for the proteins are not as good as those of the other structures, but they are still satisfactory.

Table 1 DSC measures and evaluation of the segmentation results.

Image of	DSC Mean	Std. Dev.
Nucleus	0.89	0.07
Protein	0.75	0.18
Cell	0.86	0.04

In the following, the influence of the segmentation parameters is evaluated depending on the size of the structural element for the morphological operators and the size of the texture window for the rugosity image. Results are presented in tables 2 and 3.

Table 2 Results for various texture window sizes.

Texture window size	DSC Mean	Std. Dev.
2	0.65	0.15
4	0.86	0.04
6	0.87	0.09

Table 3 Results for various structural element sizes involved in the nucleus segmentation.

Structural	Element Size	DSC Mean Std. Dev.
6	0.8835	0.0712
5	0.8757	0.0694
4	0.8870	0.0679

The same problem appeared with the nucleus segmentation, but less often. For our system, it was crucial to deal with such segmentation errors, because they could bias values of spatial features computed for characterizing cell states (see next Section). The problem mentioned above reveals that gray-level information used for thresholding operations is not sufficient for the nucleus and cell segmentations. The problem can be fixed by

combining information from successive slices and from various image channels. The idea is illustrated in fig. 11.c and fig. 11.d. The presence of several nuclei in the DAPI channel is usually well detected by the segmentation algorithm, even when corresponding cells in the RHOD channel are merged. Hence, the nucleus segmentation results can be used to check if the number of segmented cells is correct and, if necessary, to correct cell segmentation results.

The correction relies on the computation of the discriminant line which best separates the nucleus regions using a criterion of linear discriminant analysis (13). In the same manner, by checking the coherence of nucleus segmentations over the slices and, combining results of various slices, possible merged segmentation are corrected. The database does not contain cases with more than two cells but the method can be easily extended for splitting several cells (it is a generalization of the linear discriminant analysis). An automatic system for the analysis of intercellular communication and early carcinogenesis.

Spatial characterization

In this part, preliminary results are presented about the core of our global project: automatic spatial reasoning with these segmented images in the same way as biologists do. According to biological hypothesizes, pathological cell states are characterized by specific Cx43 spatial distributions within cells. Relying on biological observations on 2D slices of the cell, we compute 2D spatial features for each cell (3 slices) such as:

- Number of proteins in the image,
- Average distance from the proteins to the nucleus,
- Average distance from the proteins to the cell membrane.

The minimal distance between the protein and the membrane was computed by successive morphological erosions but the distance between the protein and the nucleus was only a point to point distance with the center of the nucleus. That questionable choice was required first for computing complexity problems but should be further investigated. Last, the values on the Y-axis are the average of the features on three successive slices of the sample

Despite an improvable cell segmentation quality, initial results allow us to point out significant differences between pathogens class and control class. Fig. 12 shows graphical results for each index indicating that the automatic computation of distances between the various structures within a cell will give efficient measures to classify the pathogens cells and the control ones. These seminal results - which are mainly numerical - encourage us to imagine a symbolical representation of these spatial characteristics in the way that biologists express themselves: for instance, "proteins are quite aligned near the cell membrane" (13). As a matter of fact, control cell sample 3 indicates that the proteins are quite close from both the nucleus and the cell membrane. But these measures are absolute measures in pixels not taking into account the relative size of the cell for instance.

DISCUSSION

The aim of this primary study was to develop a prototype of computer-assisted tool for exploring the mutual organization in 3D space of proteins, nucleus and cytoplasm within a cell. The manual classification and the ground truth for a set of 60 cell slices which were given by the biologist allowed us to validate quantitatively the designed image analysis tools. The cell structures segmentation rates achieved in this study encourage us to go further into the spatial characterization of 3D configurations of these cell structures. Preliminary statistical features were computed from a set of 11 cells belonging either to control or pathogens class (fig. 12). These measures gave the initial quantitative evidence that an automatic cell classification tool based on the qualitative spatial characterization within 3D images is possible.

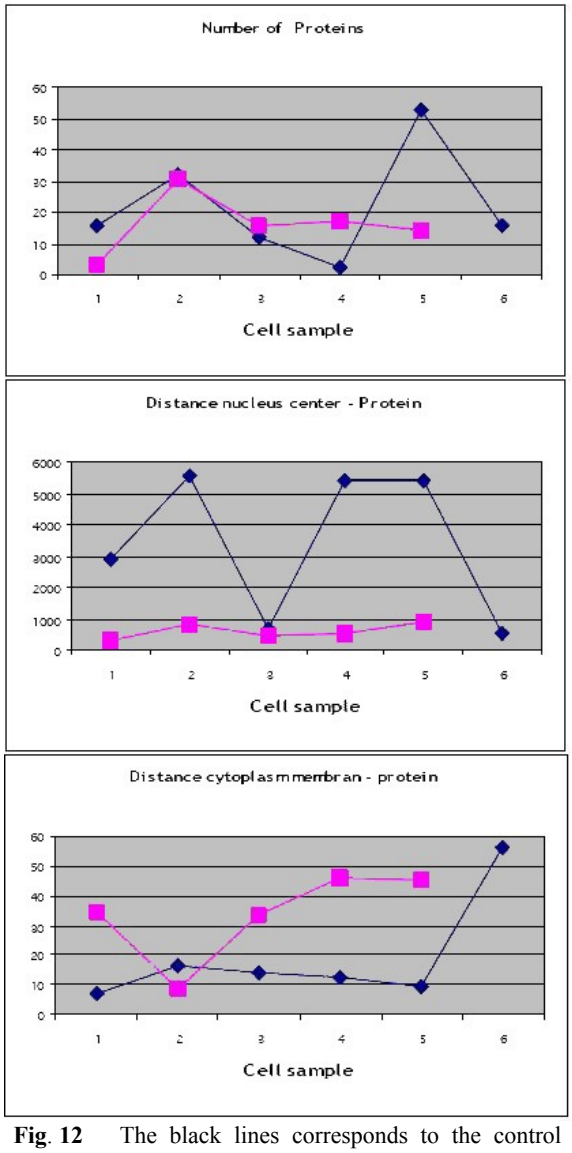


Fig. 12 The black lines corresponds to the control class and the grey lines to the pathogens class.

Computing more criteria is indeed a possible extension of this work which only gives some insights into cell classification. However, a statistical validation of criteria and feature would require a larger database we do not have yet. Here our aim was just to show the feasibility of the classification using the simplest features in terms of complexity. In the meantime, we draw some perspective points to be further investigated both technically and theoretically at the next step: develop an open source friendly user interface for the interactive exploration of the 3D cell. The computed features must be robust to low signal-to-noise ratios and hence the segmentation module must be improved by much more sophisticated methods such as the multi-scale analysis (10). Last, a research effort has to be carried out for the formal definition of valid and semantically coherent spatial relationship between the cell structures such as the notion of proximity or the very idea of distance between two structures.

REFERENCES

1 W.R. Lowenstein and B. Rose, The cell-cell channel in the control of growth. Semin. *Cell*. *Biol*., 1992; **3**: 59-79.

2 N.M. Kumar and N.B. Gilula, The gap junction communication channel, *Cell*, 1996, **84**: 381-388.

3 R.J Ruch, Role of gap junctions in cellular growth control and neoplasia evidence and mechanism. In: Peracchia C, editor. Current topics in membrane gap junctions, San Diego: *Academic Press*, 2000, **49** : 535-554.

4 H. Yamasaki, Y. Omori, V. Krutovskikh et al. Connexins in tumor suppression and cancer therapy. *Novartis Found Symp*.,1999, **219**: 241-254.

5 D. Laird, P. Fistouris, G. Batist et al. Deficiency of connexion. 43 gap junctions is an independent marker for breast cancer. *Cancer Res*, 1999, **59**: 4104-4110.

6 D. Segretain, B. Mograbi, J. Dompierre et al, Sequestration of connexin 43 in the early endosomes is an early event of leydig cell tumor progression, *Molecular carcinogenesis*, 2003, **38**: 179-187.

7 R. Duda, P. Hart and D. Stork, Pattern Classification, Ed. *Wiley-Interscience*, 2001.

8 Jean Serra, Image Analysis and Mathematical Morphology" *Ac*. *Press*, *Volume 1* (1982), *Volume 2* (1988).

9 N. Otsu., A Threshold selection method from gray level histograms. IEEE Trans. *System Man Cybernet*.,1979, **9**: 62-66.

10 J.-C. Olivo-Marin Extraction of spots in biological images using multiscale products, Pattern Recognition, Issue 9, September 2002, **35**: 1989-1996.

11 L. R. Dice, Measures of the amount of ecologic association between species, *Ecology*, 1945, no. 3, **26**: 297-302.

12 A. P. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer, Morphometric Analysis of White Matter Lesions in MR Images: Method and Validation, IEEE Transcations on Medical Imaging, *december* 1994, **13**: 716-724.

13 Jochen Renz, Qualitative Spatial Reasoning with Topological Information, Springer, *Lectures Notes in AI*, 2002.